2017

Moving beyond presence and absence when examining changes in species distributions

Michael B. Ashcroft
University of Wollongong, ashcroft@uow.edu.au

Diana H. King
University of Wollongong, dhk442@uowmail.edu.au

Ben Raymond
Australian Antarctic Division

Johanna Turnbull
University of Wollongong, jdt17@uowmail.edu.au

Jane Wasley
University of Wollongong

See next page for additional authors

Publication Details

Ashcroft, M. B., King, D. H., Raymond, B., Turnbull, J., Wasley, J. & Robinson, S. A. (2017). Moving beyond presence and absence when examining changes in species distributions. Global Change Biology, 23 (8), 2929-2940.

Moving beyond presence and absence when examining changes in species distributions

Abstract

Species distributions are often simplified to binary representations of the ranges where they are present and absent. It is then common to look for changes in these ranges as indicators of the effects of climate change, the expansion or control of invasive species or the impact of human land use changes. We argue that there are inherent problems with this approach, and more emphasis should be placed on species relative abundance rather than just presence. The sampling effort required to be confident of absence is often impractical to achieve, and estimates of species range changes based on survey data are therefore inherently sensitive to sampling intensity. Species niches estimated using presence-absence or presence-only models are broader than those for abundance and may exaggerate the viability of small marginal sink populations. We demonstrate that it is possible to transform models of predicted probability of presence to expected abundance if the sampling intensity is known. Using case studies of Antarctic mosses and temperate rain forest trees we demonstrate additional insights into biotic change that can be gained using this method. While species becoming locally extinct or colonising new areas are extreme and obviously important impacts of global environmental change, changes in abundance could still signal important changes in biological systems and be an early warning indicator of larger future changes.

Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details

Ashcroft, M. B., King, D. H., Raymond, B., Turnbull, J., Wasley, J. & Robinson, S. A. (2017). Moving beyond presence and absence when examining changes in species distributions. Global Change Biology, 23 (8), 2929-2940.

Authors

Michael B. Ashcroft, Diana H. King, Ben Raymond, Johanna Turnbull, Jane Wasley, and Sharon A. Robinson

Opinion – Global Change Biology

Moving beyond presence and absence when examining changes in species distributions

Running title: Moving beyond presence and absence

Word count: 5036 (introduction to acknowledgements inclusively; 5 Figures, 2 colour)

Ashcroft, Michael B. [1,*]

King, Diana H. [1]

Raymond, Ben [2,3]

Turnbull, Johanna D. [1]

Wasley, Jane [1,2]

Robinson, Sharon A. [1]

[1] Centre for Sustainable Ecosystem Solutions, School of Biological Sciences, University of Wollongong, Wollongong, New South Wales, 2522, Australia.

[2] Australian Antarctic Division, Department of Environment, Kingston, Tasmania, 7050, Australia.

[3] Antarctic Climate and Ecosystems Cooperative Research Centre, Private Bag 80, Hobart, Tasmania 7001, Australia.

[*] Corresponding author: ashcroft@uow.edu.au Ph. +61 4221 3454

Keywords: Antarctic mosses, climate change, geographic range, negative binomial distribution, population abundance, presence-absence data, range shifts

Abstract

Species distributions are often simplified to binary representations of the ranges where they are present and absent. It is then common to look for changes in these ranges as indicators of the effects of climate change, the expansion or control of invasive species or the impact of human land use changes. We argue that there are inherent problems with this approach, and more emphasis should be placed on species relative abundance rather than just presence. The sampling effort required to be confident of absence is often impractical to achieve, and estimates of species range changes based on survey data are therefore inherently sensitive to sampling intensity. Species niches estimated using presence-absence or presence-only models are broader than those for abundance and may exaggerate the viability of small marginal sink populations. We demonstrate that it is possible to transform models of predicted probability of presence to expected abundance if the sampling intensity is known. Using case studies of Antarctic mosses and temperate rain forest trees we demonstrate additional insights into biotic change that can be gained using this method. While species becoming locally extinct or colonising new areas are extreme and obviously important impacts of global environmental change, changes in abundance could still signal important changes in biological systems and be an early warning indicator of larger future changes.

Introduction

The distributions of species are often presented as geographic ranges, binary classifications of where they are present and absent (Gaston, 2009). These can be useful guides for where species may be encountered, but usually lack detail on variations in density and the patchiness of occurrence within these ranges. This limitation is particularly important near range boundaries, where populations may be so small, sparse and variable that it is virtually impossible to be certain where a species becomes entirely absent (Brown *et al*., 1996; Hanewinkel *et al*., 2014; Seipel *et al*., 2016). If species ranges are only used as rough guides then uncertainty on range boundaries are perhaps not critical. However, when investigating the effects of climate change on species ranges (Parmesan & Yohe, 2003; Hickling *et al*., 2005, 2006), the spread or eradication of invasive species (O'Donnell *et al*., 2012), or the impacts of changes in human land use (Ameztegui *et al*., 2016) it is crucial to determine species distributions accurately so that changes can be reliably detected. In this article we highlight some of the issues from focusing only on presence and absence of a species when examining the impacts of environmental change and suggest some alternatives that can be used, even when data is only collected in presence-only or presence-absence format.

Defining species ranges using abundance estimates

Species ranges are often determined by examining samples of communities across an environmental gradient or region. Each sample may consist of museum or herbaria records in a 10 km grid cell, a survey of 1 ha quadrats, or any other data depending on what is available. Usually studies have sufficient data to determine where the core of a species distribution is located, but if there are false absences (present but not observed) near range boundaries then there is a risk that range sizes will be underestimated and changes in range incorrectly inferred (MacKenzie et al., 2002; Bates *et al*., 2015). While false absences near range cores

may have no effect on the overall ranges, it is crucial to be certain of absences near the boundaries, as even a single observation may extend the estimate of the overall range or imply a species distribution has changed if surveys from multiple time periods are compared.

In this section we examine the degree to which we can be confident in delineating range boundaries by where species become absent. We are primarily concerned with situations where a finite community size for each grid cell or quadrat can be estimated (e.g. by extrapolating from subsamples or similar communities) and the effort that has been expended to sample each community is known. In our model, the community size is assumed to be an abundance of individuals, but similar methods could be used if community size was estimated in terms of metrics such as cover. For simplicity we assume each sampled cell or quadrat has a discrete community of individuals and populations of species, even though we acknowledge that actual communities and populations may extend outside sampled units.

To illustrate the difficulty in being confident of absence, consider a situation where there is an estimated community of 1 000 trees in a grid cell near a range margin, and the requirement is to determine whether species X is present. We assume there is no prior knowledge on whether the species is present, although situations with prior knowledge are discussed later. It is only necessary to find one individual of species X to confirm its presence, but to be absolutely certain of absence it is necessary to sample all 1 000 trees. Even if we sampled 999 trees without finding the species, it is possible the last tree would confirm presence. It therefore becomes impractical to separate complete absence from low density populations as the size of communities or sampling units becomes larger. It might be possible to reduce the effort required by only requiring a 95% certainty of species absence (Garrard *et al.*, 2014), but how much sampling effort is required to achieve that?

We addressed this problem using a Bayesian model to estimate the overall population of a species that was not observed in a sample (model included in Supporting Information).

To do this we simulated taking a sample of n individuals from a total community of size N, and estimated the 95% CI for the total population of a species (M) in the community, given the species was not observed (m = 0). While it was obviously always possible for the overall population M to be zero when the sample m was zero, if the sample size was small then the expected population (and 95% CI) could still be quite large due to the substantial possibility of non-detection.

We found that as the community size (N) increased from 10 to 100 000 (e.g. Fig. 1a, b illustrate N = 100 and 1 000 respectively) the proportion of sampling needed to confirm a low relative abundance decreased rapidly, but it was still necessary to sample 76–80% of the individuals in a community to be 95% sure a species was absent (M = 0; Fig. 1c). In practice, ascertaining whether a species is 95% likely to be absent is possible for small communities, and has been suggested as a possibility for environmental impact studies (Garrard *et al*., 2014). However, it is an impractical solution for larger communities or for examining whether ranges have changed over time if the sampled units are large areas, such as 10 km grid cells (e.g. Hickling *et al*., 2005, 2006).

An alternative approach that we advocate is to define ranges in terms of a threshold relative abundance (M / N). That is, instead of constructing ranges by determining where a species is most likely absent (e.g. 95% certainty that M = 0), which is often impractical to answer, it is possible to define them in terms of where the expected relative abundance of the species is less than an arbitrary threshold, such as M / N < 1%. For small communities (N = 10–100 individuals; Fig. 1a, c) the sampling effort required to answer this question is comparable to investigating absence, but for larger communities the sampling effort is much lower (Fig. 1b, c). For example, in a community of 1 000 individuals it is only necessary to sample 5% of the community in order to estimate a relative abundance below 1%, or 29% of the community for an abundance less than 0.1% (Fig. 1c). This is much more achievable in

practice than the 78% sample coverage required to be 95% confident of absence. The threshold can be defined such that the range only includes locations where the population abundance is sufficient to be ecologically viable, and can exclude locations where the species is only present in small marginal sink populations. Of course, the vague definitions of which communities are small and large will depend on factors such as the scale of the study and the available sampling budget, as it is a reflection of what is feasible to survey.

The model presented here assumes there is a known finite community size and that individuals can be selected at random. In this paper we assume this is acceptable for studies of trees in a forest and moss shoots in a slow growing Antarctic environment. However, there may be limitations when applying the models to mobile, cryptic or secretive species, or when sampling is biased towards roads or other features. While questions remain in these situations, we suggest that similar limitations still apply to presence-absence studies. It is still difficult to be confident of absence (e.g. Wintle *et al*., 2005), and therefore determine if species distributions have changed, and even where a species is present, it is still important to know if it has sufficient abundance to be a viable population.

Transforming probability of presence to relative abundance

In the previous section we suggested that species ranges should be defined in terms of relative abundance rather than simply presence or absence. This is a pragmatic suggestion based on sampling effort, and is particularly relevant to determining past changes in species ranges by repeated sampling of standardised quadrats or transects. However, in many cases the only data available are in presence-absence format, or even presence-only (Elith & Leathwick, 2009; Phillips *et al*., 2009), and we are often interested in modelling changes in species ranges. Indeed, in many cases changes in species ranges are inferred by ecological niche (Pearson & Dawson, 2003; Araújo & Peterson, 2012; Bates *et al*., 2015) or dynamic

models (Schurr *et al*., 2012; Merow *et al*., 2014; Pagel *et al*., 2014) rather than intensive sampling *per se*. This leads to two further questions: How do species response curves for presence compare to those for abundance, and is it possible to estimate abundance from commonly used models of probability of presence?

We addressed these questions by simulating the abundance of a species population (M) using a negative binomial model (O'Hara & Kotze, 2010; Lindén & Mäntyniemi, 2011; Warton *et al*., 2016) and then once again randomly sampling n individuals from an overall community of size N using a hypergeometric distribution. By repeating this 10 000 times we were able to examine how the probability of occurrence in the sample, $P(m > 0)$, related to the expected relative abundance of the species in the overall community (M / N). Note that unlike the previous section both probability of presence and expected abundance are threshold independent, although a threshold could be applied later to convert these to a binary range.

The negative binomial distribution used to simulate the population abundance (M) can take different shapes, approximating Gaussian (Fig. 2b) or exponential decay (Fig. 2a) distributions depending on the parameterisations. As it is not possible for the species population (M) to be greater than the total community size (N) the curves were truncated and scaled so that the total probability summed to one. The distribution of the abundance of the species in the sample (m; Fig. 2c, d) had a similar shape to that of the overall population (M) but was zero inflated. That is, there was more likelihood of zero or low abundance in the sample due to the possibility the species was present but not sampled.

The negative binomial model was parameterised in R (R Core Team, 2015) using two parameters: mu representing the mean population, and shape (or size; r) affecting the variability or dispersion (models included in Supporting Material). We assumed the shape parameter was constant for a species and that mu could change according to environmental

suitability along a gradient. By varying mu and keeping the community size (N), sample size (n) and shape parameter (r) constant we were able to examine how the probability of sample presence related to the overall expected size of the species population as a relative abundance within the overall community.

We found that there were non-linear relationships between the probability of presence in the sample and expected relative abundance in the overall community. The estimated relative abundance (M / N) increased very slowly at first, until the probability of sample presence was approximately 40%–80%, and then started to increase more rapidly. As the sampling proportion (n / N) or shape parameter (r) increased this non-linear effect became stronger such that even higher probabilities of sample presence were possible even if the actual expected populations were relatively small (Fig. 2e, f).

This represents a potential hazard for conservation planning or change detection. If the probability of presence can be high even when populations are relatively small, presence-absence models might exaggerate the niche width, potentially causing conservation actions to be directed towards small marginal sink populations, which rely on dispersal from other locations to maintain viable populations, rather than high abundance core populations which are self-sustaining in their own right. To illustrate this, consider a dataset of three mosses (*Schistidium antarctici* (Card.) L. Savic. & Smirn. (syn. *Grimmia antarctici* Card.), *Bryum pseudotriquetrum* (Hedw.) Gaertn., B. Mey. & Scherb. and *Ceratodon purpureus* (Hedw.) Brid.) that occur along a moisture gradient near Casey Station in Antarctica. *Schistidium antarctici* occurs at the wetter end of the gradient, with a sparser mixture of *C. purpureus* and moribund moss (dead or dying unidentifiable mix of moss species) at the drier end and *B. pseudotriquetrum* occurring to some extent in both communities.

We had presence-absence data for the three species and moribund moss (hereafter simplified to four ‘species’) from 60 quadrats along a water availability gradient, which we

collected in 2003 (sites and quadrats detailed in Wasley *et al*., 2012; Ashcroft *et al*., 2016). We initially modelled the presence of each species using a binomial generalised additive model (GAM; Hastie & Tibshirani, 1990). We then transformed the predicted probability of presence to the estimated relative abundance along the moisture gradient for each species using models like those in Fig. 2e. We assumed the sample size was approximately 200 moss shoots (nine pinches of up to 20–50 shoots with mosses not always present; Wasley *et al*., 2012) from a total community of 260 000 (pinches taken in 20 cm × 20 cm quadrats with densities ~ 650 shoots/cm^2; Wasley *et al*., 2006; sampling intensity is intentionally low to minimise impact). The shape parameter was estimated by ensuring the total abundance of all four species was approximately 100%, and we found a shape of r = 0.4 gave a good approximation at the moist end of the gradient where moss cover is high (Fig. 3c). The total abundance is expected to be less than 100% at the dry end of the gradient as lichens become more abundant and displace mosses (Melick and Seppelt, 1997).

As expected, we found that the niches for all species appeared broader in terms of predicted probability of occurrence (Fig. 3a) than in terms of predicted abundance in the community (Fig. 3b). As we suggested based on Fig. 2e, f, this is due to the possibility that a species can have a high probability of occurrence even when the relative abundance is low. The binomial models suggested that all four species were likely to be present along the entire moisture gradient, although the transformed abundance models suggested the abundance of some species was very low at the extremes which better matches observations (Selkirk & Seppelt, 1987; Wasley *et al*., 2012).

To validate predicted abundances we used photos of 40 of the quadrats obtained at the same time the samples were taken. We calculated the cover of live and moribund moss from these photos (D.H.K. unpublished data) and then used the estimated proportions of each species in each quadrat (based on the Braun–Blanquet methods in Wasley *et al*., 2012) to

divide the cover of live moss into the three individual species. We found the relationships between the predicted probability of presence and cover (Fig. 4a, c, e, g, i) all resembled the theoretical non-linear curves in Fig. 2e, f where the cover generally remained low until the probability of occurrence exceeded approximately 80%. We found that our predicted abundance was a better predictor of cover, with improved correlations and much less bias compared with the presence-absence predictions (Fig. 4b, d, f, h, j).

As a further example, consider the distribution of rainforest trees on the Illawarra Escarpment, 80km south of Sydney, Australia (Ashcroft *et al*., 2008, 2009, 2012a). Presence absence data for 21 rainforest species were collected from 600 20 m × 20 m quadrats and individual species distributions were modelled using Generalised Additive Models and fine scale grids of geology, winter minimum temperature, summer maximum temperature and summer minimum temperature (full details on species and environmental layers in Ashcroft *et al*., 2008, 2012a). Landscape scale variations in warming have been estimated by relating microclimatic observations with Bureau of Meteorology data (Ashcroft *et al*., 2009), with future conditions (2040) estimated by assuming current trends continue (thus adding past change from 1972 to 2006 onto 2006 conditions).

The richness of rainforest species was estimated by summing the probabilities of presence for the 21 species (D'Amen *et al*., 2015). This suggested rainforest species were present across much of the study area (Fig. 5), and while this is true, they are often only present in the understorey of communities of *Eucalyptus* spp., *Acacia* spp. and other species. The actual rainforest communities, where rainforest species have a higher relative abundance, are restricted to locations that are sheltered from the hot, dry north-westerly winds that can be desiccating for rainforest species and have a large effect on maximum temperatures (NPWS, 2002; Ashcroft *et al*., 2008).

We again converted the probability of presence of each rainforest species to expected abundance by estimating the sampling intensity of each quadrat (n = 95, N =100) and ensuring the sum of abundances was approximately 100% in the rainforest patches (r = 2; See Figure S1 in Supporting Information for sensitivity analysis with other values of r). We found that locations where rainforest species were predicted to be abundant were much more restricted in distribution than the areas where the rainforest species were predicted to be species rich (Fig. 5). This provides further support that presence-absence niches are broader than the niches for species abundance. Furthermore, the predicted future changes in abundance were quite poorly correlated with predicted changes in richness ($r^2 = 0.299$). The hot, dry inland areas in the northwest were predicted to decline in both richness and abundance. The moistest, sheltered rainforest patches along the escarpment were predicted to increase in both abundance and richness. However there were also areas where abundance and richness exhibited opposite trends, which may occur if the community is shifting to or from a community where a few species are abundant but richness is low. The overall trend towards increasing richness along the escarpment may signal a shift from Coachwood Warm Temperate Rainforest to the higher diversity Illawarra Escarpment Subtropical Rainforest (NPWS 2002).

While there are obviously assumptions in these models (e.g. estimated sampling intensity and shape parameters may vary between species and locations) the results are qualitatively similar if these values are varied (e.g. Figure S1). The estimated niche width and distribution of species are still broader based on presence-absence models than they are based on abundance models. Presence-absence models may therefore exaggerate the niche widths or distribution of species and place undue emphasis on small marginal populations (similar to the naughty noughts problem discussed by Austin & Meyers, 1996). Our results demonstrate that it is possible to predict abundance from probability of presence, if the

sampling intensity and shape parameter of the negative binomial distribution can be estimated. We therefore suggest that converting probability of presence to the expected population size allows new ecological interpretations of species niches and ranges that are not possible with presence-absence predictions alone.

Models for probability of presence are often produced using presence-only data, from which it can be difficult to estimate the sampling intensity, species prevalence, or sampling biases (Elith *et al*., 2011). It is not yet clear how to best estimate these parameters, however, while the transformation curves can vary considerably (Fig. 2e, f), we have demonstrated with our examples that approximate estimates are sufficient to generate good predictions of abundance (Fig. 4) and vastly different distributional patterns (Fig. 5, S1).

Dealing with the limitations of uncertain absences

We have argued that it is often impractical to prove absence, and ecologists should be focusing on relative abundance rather than simply presence. We have demonstrated that Bayesian models can be used to estimate abundance even when a species is not observed (Fig. 1), and that models for probability of presence can be converted to expected abundance if the sampling intensity and the shape of the negative binomial distribution can be estimated (Fig. 3–5). This does not resolve all problems, however, because many ecological concepts and studies rely on establishing the absence of a species. In this section we discuss methods to deal with the limitation of uncertain absences in a global change context.

Range changes

Species ranges have attracted particular attention in climate change literature (Parmesan & Yohe, 2003) and yet there are problems with determining ranges or range changes. For example, range boundaries are known to be sensitive to sampling intensity, so that greater

sampling in one time period can lead to apparent range shifts (Hassall & Thompson, 2010). Numerous methods have been developed to address this, including sub-sampling or examining the mean or median location of species records (Shoo *et al*., 2006; Maclean *et al*., 2008; Hassall & Thompson, 2010). Importantly, these corrections do not improve the accuracy with which range boundaries are identified, but change the focus to look more at the core distribution or less extreme margins.

Hanewinkel *et al*. (2014) divided the range of a species into the core, extended and occasional occurrence areas based on the likelihood of observing the species. While we have argued that the limits of the occasional occurrence area are generally impractical to delineate, greater recognition of the issue could lead to more meaningful and accurate studies on changes in the core or extended distributions. Rather than claim to be examining changes in a species overall range (including occasional occurrences), scientists could define clear criteria for the extended range boundary in terms of a threshold probability of occurrence, or as we have argued, a minimum relative abundance.

There is no reason to believe the range boundaries will shift in the same manner as the core distribution. In fact, it is recognised that different processes operate in different portions of a species range (Hampe & Petit, 2005). The leading edge is likely to be influenced by dispersal ability, the trailing edge may be characterised by population fragmentation and potentially long-term persistence in microrefugia before the overall range actually contracts, and the core distribution may undergo a crash in population long before the probability of occurrence begins to decline substantially (Fig. 2e, f). In all cases, focusing on abundance is likely to give a different picture of climate change impacts than simply presence or absence (Fig. 5). Studies of range shifts that focus on means or centroids of ranges (e.g. Shoo *et al*., 2006; Maclean *et al*., 2008) are likely to be focused on the core distribution, while those based on the average of extreme observations (e.g. Hickling *et al*., 2005, 2006) are more

likely to detect changes in the extended or occasional areas depending on sampling intensity. The results are not necessarily comparable in meta-analyses unless differences in sampling methodologies are explicitly catered for (Brown *et al*., 2016).

Regardless of whether climate change impacts are assessed using abundance or presence, it is difficult to be confident of range shifts if populations are variable (McCain *et al*., 2016). Studies that are based on too few points in time are likely to overestimate change when they confound long-term trends with short-term variability (Brown *et al*., 2016). While there is no easy solution to this, it is possible to minimise the component of short-term variability that is due to sampling intensity if less emphasis is placed on the extreme range margins that are inherently difficult to establish accurately.

Prior knowledge and environmental impact assessment in a changing environment

Biodiversity can be impacted by changes in land use, and legislative approval of human activities can be impacted by the presence or absence of threatened species (Garrard *et al*., 2014). Recently, there has been increasing recognition that detectability is not perfect and the probability of observing presence can depend on factors such as sampling intensity and the cryptic nature of species (Chen *et al*., 2009). If detectability and occupancy are affected by different factors then potentially the two can be separated (MacKenzie *et al*., 2002). However, if abundance varies along an environmental gradient, then changes in abundance can potentially affect both occupancy and detectability making it more difficult to separate the two factors (Wintle *et al*., 2005). Detectability and occupancy are likely to decrease simultaneously as populations become smaller and more isolated.

A further issue is that if it is accepted that it is only possible to be 95% sure of absence (e.g. Garrard et al. 2014), this requires a value judgement to be made. Is it necessary to be 95% sure of absence to allow development to proceed, or 95% sure of presence to halt

activities? Is it possible to quantify the impact to threatened species simply by their presence or absence, or is it necessary to know that the species is abundant and persistent enough to form a viable population rather than just dispersing individuals or a sink population?

Garrard *et al*. (2014) suggested setting a minimum survey effort based on the prior expectation of species presence at a site. A similar approach could be used for abundance. Given a prior expectation of species abundance at a site (e.g. from models such as Fig. 3b, 5) it is possible to estimate the survey effort required to develop a suitable estimate of abundance. Even if only presence-absence data are collected, a Bayesian model can still be used to refine the estimate of abundance (e.g. Fig. 1; Royle & Nichols, 2003), and quantifying the population abundance avoids the value judgement discussed above. Instead of quantifying the probability the species is present or absent, this could establish whether or not the expected population is large enough to be viable. If the prior expectation is that a large population inhabits the site, a higher sampling intensity would be needed to confidently confirm a low estimate for the population. If we are confident the species is absent then a lower sampling effort is sufficient.

Changes in endemism

Unless we are dealing with a well-known species (see discussion of prior knowledge in previous section), to experimentally prove a species is endemic to one region it is necessary to prove it is absent from all others. It may be trivial to establish that, for example, giraffes are absent from Antarctica or even endemic to parts of Africa, but for the majority of species, which are more difficult to detect and have less prior knowledge, it is very difficult to make such conclusions. Most species (e.g. invertebrates) have not been described, their ranges quantified, or the environmental determinants of their distributions identified (the Linnean, Wallacean and Hutchinsonian shortfalls, Whittaker *et al*., 2005; Mokany & Ferrier, 2011).

There is therefore a risk that the concept of endemism is limited to easily-detected, charismatic or well-studied species, or regions where there is a clear barrier to dispersal so that absence can be confidently inferred elsewhere. If we cannot prove a species is absent in all other regions, it is necessary to rely on prior knowledge rather than experimental evidence.

One potential solution to this problem is to focus on multi-species and multi-scale indices of endemism, for example classifying regions where endemism is predicted to be high to avoid the burden of proving any particular species is absent elsewhere (e.g. Laffan & Crisp, 2003). This makes intuitive sense in conservation where it is more important to identify hot spots where many endemic species are threatened (Myers *et al*., 2000) than to conclusively establish the endemism of particular species. From a change perspective we can then examine changes in collective properties of biodiversity like endemism and richness (Fig. 5) rather than focus on individual species.

Rare species

In some cases it is particularly important to separate low abundance populations from actual absence. For example, it may be important to eradicate an unwanted species (Regan *et al*., 2006), determine where a newly introduced species has colonised (Ashcroft *et al*., 2012b), distinguish between species absence in northern Europe during the last glacial maximum or persistence in microrefugia (Stewart & Lister, 2001), focus on the distribution of rare species (Engler *et al*., 2004), or confirm if a species is extinct (Reed, 1996). In all these cases the species of interest is likely to be present in low relative abundances. While it still may not be possible to be certain of absence in these circumstances, the threshold relative abundance probably needs be set lower than usual. This will increase the sampling intensity required but minimise the potential population size if the species is present (Fig. 1). This then becomes an

389 economic question of how much sampling effort is justified given the expected benefits or
390 risks (Regan *et al*., 2006).
391
392 **Conclusions**
393 Any study on changes in species ranges or distributions depends just as much on proving
394 where they are absent as it does on proving where they are present. While it is crucial to
395 distinguish between absence and low density populations when dealing with extinction, pest
396 eradication, endemism and many other ecological phenomena, the sampling effort required to
397 establish complete absence is often impractical. From a global change perspective there is a
398 lot of interest in range changes of species, but it is hard to determine exactly where species
399 become absent, let alone how ranges change over time. Our simulations and field data show
400 there can be large changes in relative abundance with very little change in presence, so
401 population crashes may occur long before change can be detected in overall ranges. While
402 much ecological data is collected in presence-absence format, and it is simpler to think of
403 species ranges in such a binary format, we have shown that it is possible to estimate
404 abundance if we can estimate sampling intensity and the shape of the negative binomial
405 distribution. This is not to say our method of analysis is the only way of dealing with the
406 issues, and indeed dynamic models of vegetation demographics may be a better approach to
407 deal with biotic interactions and disturbances if sufficient information is available to
408 parameterise such a model (Schurr *et al*., 2012; Merow *et al*., 2014; Pagel *et al*., 2014). The
409 point we stress is that there are inherent problems with any presence-absence analysis that are
410 caused by the inherent uncertainty in determining species absence. We have demonstrated
411 that it is not only simple to convert presence-absence predictions to abundance but that this
412 provides new insights into global change.
413

Acknowledgements

This research was primarily funded through Australian Research Council DP110101714 and Australian Antarctic Science grants AAS1313 and AAS4046 with logistical support from the Australian Antarctic Division for visits to Casey Station. DHK was supported by an Australian Postgraduate Award.

References

Ameztegui A, Coll L, Brotons L, Ninot JM (2016) Land-use legacies rather than climate change are driving the recent upward shift of the mountain tree line in the Pyrenees. *Global Ecology & Biogeography*, **25**, 263–273.

Araújo MB, Peterson AT (2012) Uses and misuses of bioclimatic envelope modelling. *Ecology*, **93**, 1527–1539.

Ashcroft MB, Chisholm LA, French KO (2008) The effect of exposure on landscape scale soil surface temperatures and species distribution models. *Landscape Ecology*, **23**, 211–225.

Ashcroft MB, Chisholm LA, French KO (2009) Climate change at the landscape scale: predicting fine-grained spatial heterogeneity in warming and potential refugia for vegetation. *Global Change Biology*, **15**, 656–667.

Ashcroft MB, French KO, Chisholm LA (2012a) A simple post-hoc method to add spatial context to predictive species distribution models. *Ecological Modelling*, **228**, 17–26.

Ashcroft MB, Gollan JR, Batley M (2012b) Combining citizen science, bioclimatic envelope models and observed habitat preferences to determine the distribution of an inconspicuous, recently detected introduced bee (*Halictus smaragdulus* Vachal Hymenoptera: Halictidae) in Australia. *Biological Invasions*, **14**, 515–527.

438 Ashcroft MB, Casanova-Katny A, Mengersen K, Rosenstiel TN, Turnbull JD, Wasley J,
439 Waterman MJ, Zúñiga GE, Robinson SA (2016) Bayesian methods for comparing
440 species physiological and ecological response curves. *Ecological Informatics*, **34**, 35–
441 43.
442 Austin MP, Meyers JA (1996) Current approaches to modelling the environmental niche of
443 eucalypts: implication for management of forest biodiversity. *Forest Ecology and*
444 *Management*, **85**, 95–106.
445 Bates AE, Bird TJ, Stuart-Smith RD, Wernberg T, Sunday JM, Barrett NS, Edgar GJ, Frusher
446 S, Hobday AJ, Pecl GT, Smale DA, McCarthy M (2015) Distinguishing geographical
447 range shifts from artefacts of detectability and sampling effort. *Diversity and*
448 *Distributions*, **21**, 13–22.
449 Brown JH, Stevens GC, Kaufman DM (1996) The geographic range: size, shape, boundaries,
450 and internal structure. *Annual Review of Ecology and Systematics*, **27**, 597–623.
451 Brown CJ, O'Connor MI, Poloczanska ES, Schoeman DS, Buckley LB, Burrows MT, Duarte
452 CM, Halpern BS, Pandolfi JM, Parmesan C, Richardson AJ (2016) Ecological and
453 methodological drivers of species' distribution and phenology responses to climate
454 change. *Global Change Biology*, **22**, 1548–1560.
455 Chen G, Kéry M, Zhang J, Ma K (2009) Factors affecting detection probability in plant
456 distribution studies. *Journal of Ecology*, **97**, 1383–1389.
457 D'Amen M, Dubois A, Fernandes RF, Pottier J, Pellisier L, Guisan A (2015) Using species
458 richness and functional traits predictions to constrain assemblage predictions from
459 stacked species distribution models. *Journal of Biogeography*, **42**, 1255–1266.
460 Elith J, Leathwick J (2009) Species distribution models: ecological explanation and
461 prediction across space and time. *Annual Review of Ecology, Evolution and*
462 *Systematics*, **40**, 677–97.

463 Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of
464 MaxEnt for ecologists. Diversity and Distributions, 17, 43–57.
465 Engler R, Guisan A, Rechsteiner L (2004) An improved approach for predicting the
466 distribution of rare and endangered species from occurrence and pseudo-absence data.
467 *Journal of Applied Ecology*, **41**, 263–274.
468 Garrard GE, Bekessy SA, McCarthy MA, Wintle BA (2014) Incorporating detectability of
469 threatened species into environmental impact assessment. *Conservation Biology*, **29**,
470 216–225.
471 Gaston KJ (2009) Geographic range limits of species. *Proceeding of the Royal Society B*,
472 **276**, 1391–1393.
473 Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge
474 matters. *Ecology Letters*, **8**, 461–467.
475 Hanewinkel M, Cullmann DA, Michiels H-G, Kändler G (2014) Converting probabilistic tree
476 species range shift projections into meaningful classes for management. *Journal of*
477 *Environmental Management*, **134**, 153–165.
478 Hassall C, Thompson DJ (2010) Accounting for recorder effort in the detection of range
479 shifts from historical data. *Methods in Ecology and Evolution*, **1**, 343–350.
480 Hastie TJ, Tibshirani RJ (1990) *Generalized Additive Models*. Chapman and Hall, London.
481 Hickling R, Roy DB, Hill JK, Thomas CD (2005) A northward shift of range margins in
482 British Odonata. *Global Change Biology*, **11**, 502–506
483 Hickling R, Roy DB, Hill JK, Fox R, Thomas CD (2006) The distributions of a wide range of
484 taxonomic groups are expanding polewards. *Global Change Biology*, **12**, 450–455.
485 Laffan SW, Crisp MD (2003) Assessing endemism at multiple spatial scales, with an
486 example from the Australian vascular flora. *Journal of Biogeography*, **30**, 511–520.

Lindén A, Mäntyniemi S (2011) Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, **92**, 1414–1421.

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

Maclean IMD, Austin GE, Rehfisch MM, Blew J, Crowe O, Delany S, Devos K, Decuninck B, Günther K, Laursen K, van Roomen M, Wahl J (2008) Climate change causes rapid changes in the distribution and site abundance of birds in winter. *Global Change Biology*, **14**, 2489–2500.

McCain C, Szewczyk T, Knight KB (2016) Population variability complicates the accurate detection of climate change responses. *Global Change Biology*, **22**, 2081–2093.

Melick DR, Seppelt RD (1997) Vegetation patterns in relation to climatic and endogenous changes in Wilkes Land, continental Antarctica. *Journal of Ecology*, **85**, 43–56.

Merow C, Latimer AM, Wilson AM, McMahon SM, Rebelo AG, Silander Jr JA (2014) On using integral projection models to generate demographically driven predictions of species' distributions: development and validation using sparse data. *Ecography*, **37**, 1167–1183.

Mokany K, Ferrier S (2011) Predicting impacts of climate change on biodiversity: a role for semi-mechanistic community-level modelling. *Diversity and Distributions*, **17**, 374–380.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent, J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

NPWS (2002) *Native Vegetation of the Illawarra Escarpment and Coastal Plain*. NSW National Parks and Wildlife Service, Sydney.

O'Donnell J, Gallagher RV, Wilson PD, Downey PO, Hughes L, Leishman MR (2012) Invasion hotspots for non-native plants in Australia under current and future climates. *Global Change Biology*, **18**, 617–629.

O'Hara RB, Kotze DJ (2010) Do not log-transform count data. *Methods in Ecology and Evolution*, **1**, 118–122.

Pagel J, Anderson BJ, O'Hara RB, Cramer W, Fox R, Jeltsch F, Roy DB, Thomas CD, Schurr FM (2014) Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records. *Methods in Ecology and Evolution*, **5**, 751–760.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ [accessed 24 May 2016].

Reed JM (1996) Using statistical probability to increase confidence of inferring species extinction. *Conservation Biology*, **10**, 1283–1285.

Regan TJ, McCarthy MA, Baxter PWJ, Panetta FD, Possingham HP (2006) Optimal eradication: when to stop looking for an invasive plant. *Ecology Letters*, **9**, 759–766.

535 Schurr FM, Pagel J, Cabral JS, Groeneveld J, Bykova O, O'Hara RB, Hartig F, Kissling WD,
536 Linder HP, Midgley GF, Schröder B, Singer A, Zimmermann NE (2012) How to
537 understand species' niches and range dynamics: a demographic research agenda for
538 biogeography. *Journal of Biogeography*, **39**, 2146–2162.
539 Seipel T, Alexander JM, Edwards PJ, Kueffer C (2016) Range limits and population
540 dynamics of non-native plants spreading along elevation gradients. *Perspectives in*
541 *Plant Ecology, Evolution and Systematics*, **20**, 46–55.
542 Selkirk PM, Seppelt RD (1987) Species distribution within a moss bed in greater Antarctica.
543 *Symposia Biologica Hungarica*, **35**, 279–284.
544 Shoo LP, Williams SE, Hero J-M (2006) Detecting climate change induced range shifts:
545 where and how should we be looking? *Austral Ecology*, **31**, 22–29.
546 Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota.
547 *Trends in Ecology and Evolution*, **16**, 608–613.
548 Warton DI, Lyons M, Stoklosa J, Ives AR (2016) Three points to consider when choosing a
549 LM or GLM test for count data. *Methods in Ecology and Evolution*, in press. DOI:
550 10.1111/2041-210X.12552.
551 Wasley J, Robinson SA, Lovelock CE, Popp M (2006) Some like it wet—biological
552 characteristics underpinning tolerance of extreme water stress events in Antarctic
553 bryophytes. *Functional Plant Biology*, **33**, 443–455
554 Wasley J, Robinson SA, Turnbull JD, King DH, Wanek W, Popp M (2012) Bryophyte
555 species composition over moisture gradients in the Windmill Islands, East Antarctica:
556 development of a baseline for monitoring climate change impacts *Biodiversity*, **13**,
557 257–264.
558 Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM, Willis KJ (2005) Conservation
559 biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3–23.

560 Wintle BA, Kavanagh RP, McCarthy MA, Burgman MA (2005) Estimating and dealing with
561 detectability in occupancy surveys for forest owls and arboreal marsupials. *Journal of*
562 *Wildlife Management*, **69**, 905–917.

Supporting information

BayesianHyperModel.txt – The R2OpenBugs model used to produce Fig. 1

EstimatePopulationOfUnobservedSpecies.txt – The R code that uses the model above

SimulateRandomSampling.txt – The R code to generate the data for Fig. 2–4

ProduceFig2Graphs.txt – The R code to produce the graphs in Fig. 2

cwc2003data.csv – The moss data used in Fig. 3, 4

Model2003Bryophytes.txt – The R code used to produce Fig. 3 & 4

FigureS1.tif – A sensitivity analysis for Fig. 5

571

572 **FIGURE CAPTIONS**

573 **Fig. 1.** A Bayesian model based on a hypergeometric distribution was used to predict the
574 relative abundance of a species (M / N) if n random samples were taken from a community of
575 N individuals (N = (a)100 and (b)1000) without encountering the species (m = 0). If a small
576 proportion of a community was sampled then the predicted population could be quite high
577 and uncertain even though the species was not observed. With increasing sampling intensity
578 (n / N) the expected population approached zero, but to be 95% sure the species was
579 completely absent 76–80% of the community had to be sampled regardless of community
580 size (c). In large communities it is more practical to conclude the relative abundance is below
581 some arbitrary threshold than to infer absence with any confidence.

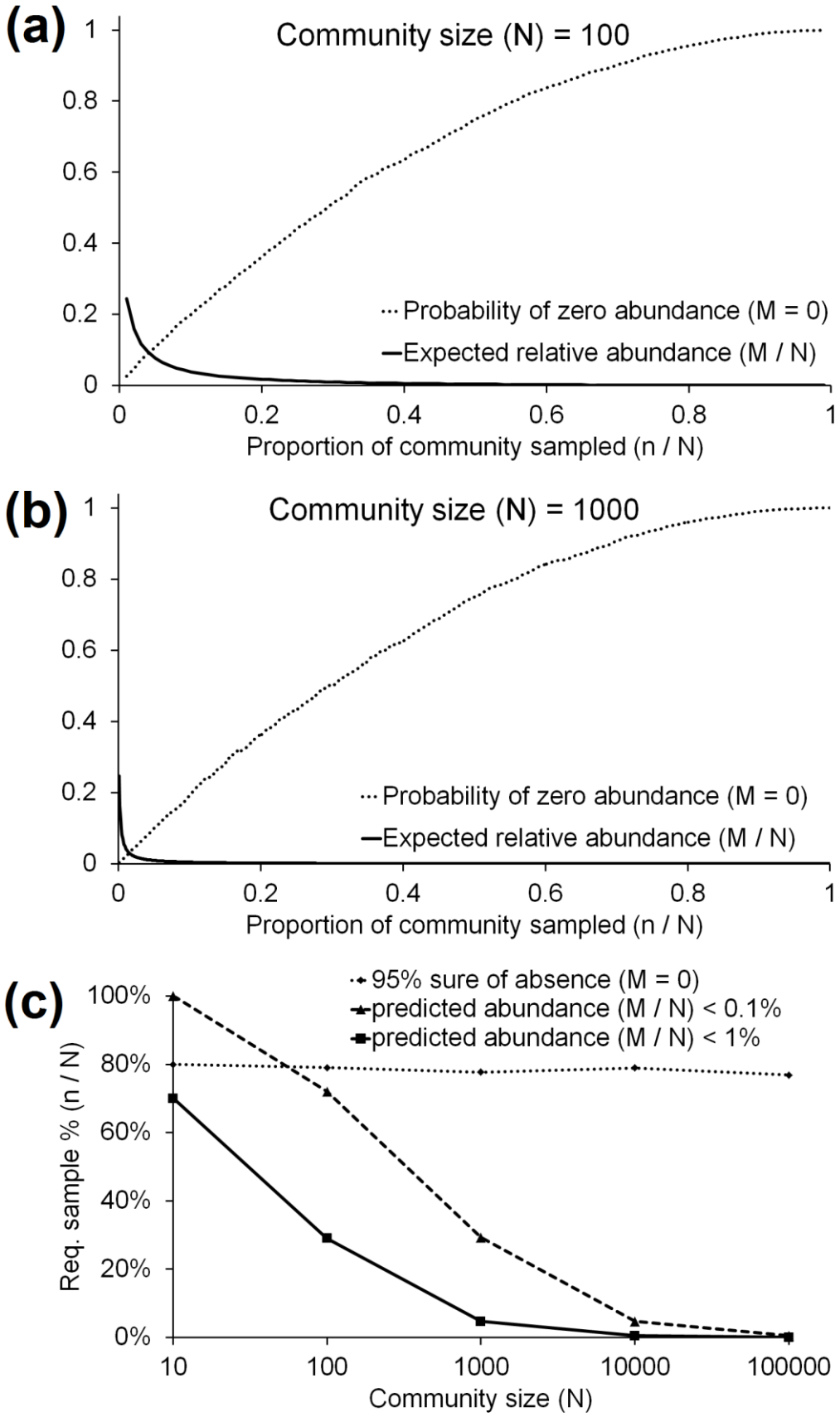
582

583 **Fig. 2.** The population size of a species (M) was simulated using a negative binomial model
584 with a community size (N) of 260 000 individuals, a shape parameter (r) of 0.4 (a) or 5 (b)
585 and selected values of mu. We then randomly sampled n = 200 individuals using a
586 hypergeometric distribution resulting in a zero-inflated distribution for the species in the
587 sample (m / n in c & d). We then examined how the probability of species presence in the
588 sample ($P(m > 0)$) related to expected relative abundance of the species in the community (M
589 / N) under different shape parameters and sampling intensities (e, f). Points in e & f
590 correspond with the different values of mu, with a curve fitted using a generalised additive
591 model on logit transformed variables (see Supporting Material).

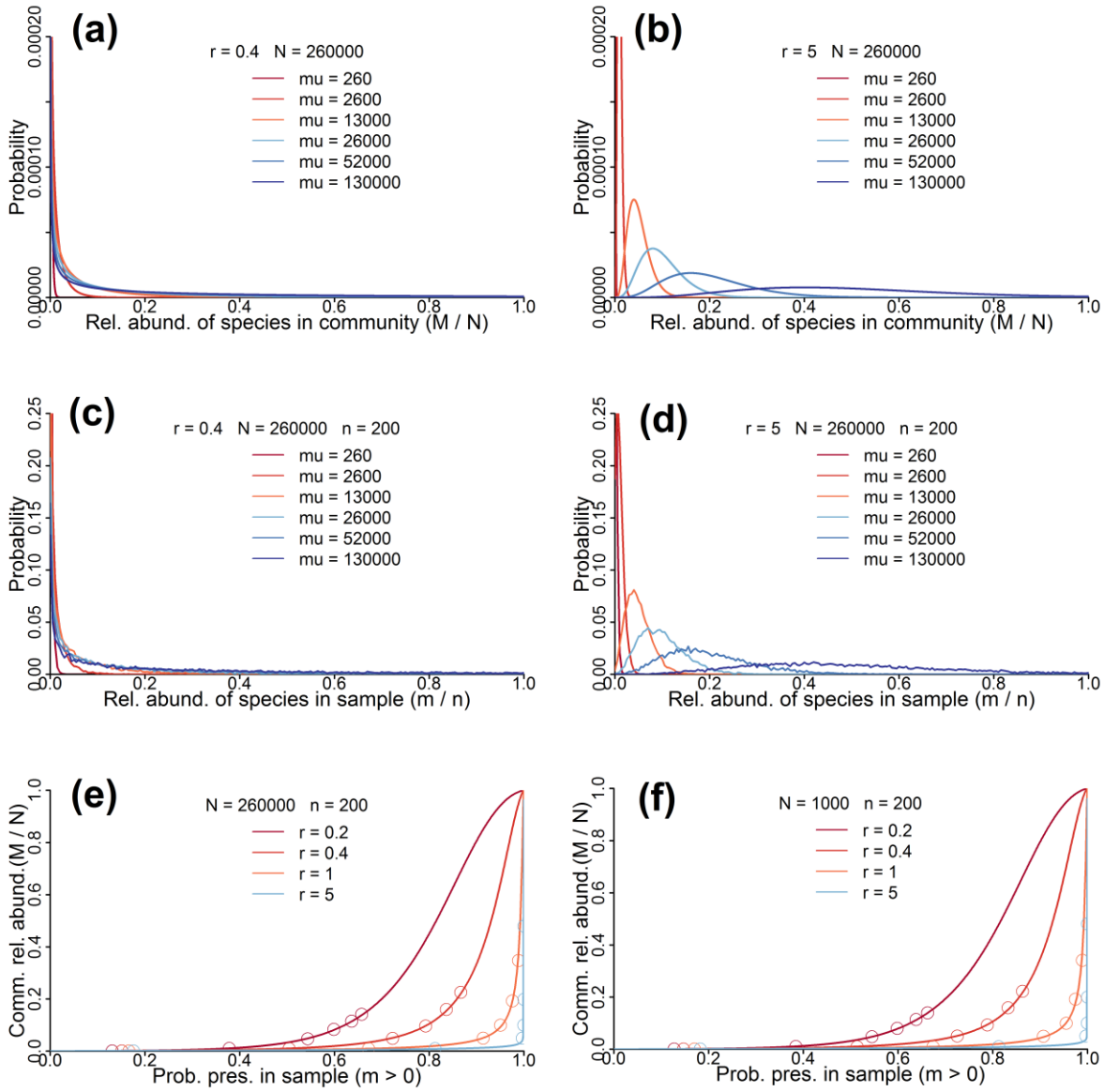
592

593 **Fig. 3.** The response of 4 moss 'species' to a water availability gradient (LogCWC; Log of
594 community water content) were examined by producing binomial generalised additive
595 models using the presence or absence in 60 quadrats (a). We then estimated the relative

596 abundance for each species (b) by assuming we had sampled 200 moss shoots from a total
597 community of 260 000 and species populations were from a negative binomial model with
598 shape (r) parameter of 0.4 (see Fig. 2) such that the total abundance of all four species was
599 approximately equal to 100% at the moist end of the gradient (c). The niche for each species
600 appeared broader in terms of presence (a) than abundance (b).
601
602 **Fig. 4.** The predictions of moss presence (left column) and abundance (right column) were
603 validated using the cover of live (a-b) and moribund (c-d) moss in photos of 40 quadrats. The
604 cover of individual species (e-j) was estimated using the proportion of each species in each
605 quadrat (based on the Braun-Blanquet methodology of Wasley *et al*., 2012) multiplied by the
606 cover of live moss. The transformed models of abundance (Fig. 3b, right column) were better
607 correlated with cover and were less biased than the presence models (Fig. 3a, left column).
608
609 **Fig. 5.** Spatial predictions of species richness and total relative abundance of 21 common
610 rainforest tree species on the Illawarra Escarpment, Wollongong, Australia (34.4 °S, 150.8
611 °E). The escarpment runs northeast to southwest through the study area and separates the
612 lowland coastal plain in the southeast from the inland Woronora plateau in the northwest.
613 Species distributions were built using Generalised Additive Models based on presence or
614 absence at 600 sample sites and grids of environmental layers (Ashcroft *et al*., 2008, 2012).
615 Estimated richness is the sum of the predicted probabilities of presence for the 21 species.
616 Probabilities of presence were converted to abundance using curves similar to Fig. 2e, f (N=
617 100, n=95, r=2). Current conditions reflect those measured in 2004 to 2006 (Ashcroft *et al*.,
618 2008). Future climates (2040) were estimated by assuming current trends continue and thus
619 we added the predicted landscape scale change from 1972 to 2006 (Ashcroft *et al*., 2009)
620 onto current conditions.


(a)
Community size (N) = 100
Probability of zero abundance (M = 0)
Expected relative abundance (M / N)
Proportion of community sampled (n / N)
(b)
Community size (N) = 1000
Probability of zero abundance (M = 0)
Expected relative abundance (M / N)
Proportion of community sampled (n / N)
(c)
95% sure of absence (M = 0)
predicted abundance (M / N) < 0.1%
predicted abundance (M / N) < 1%
Req. sample % (n / N)
Community size (N)

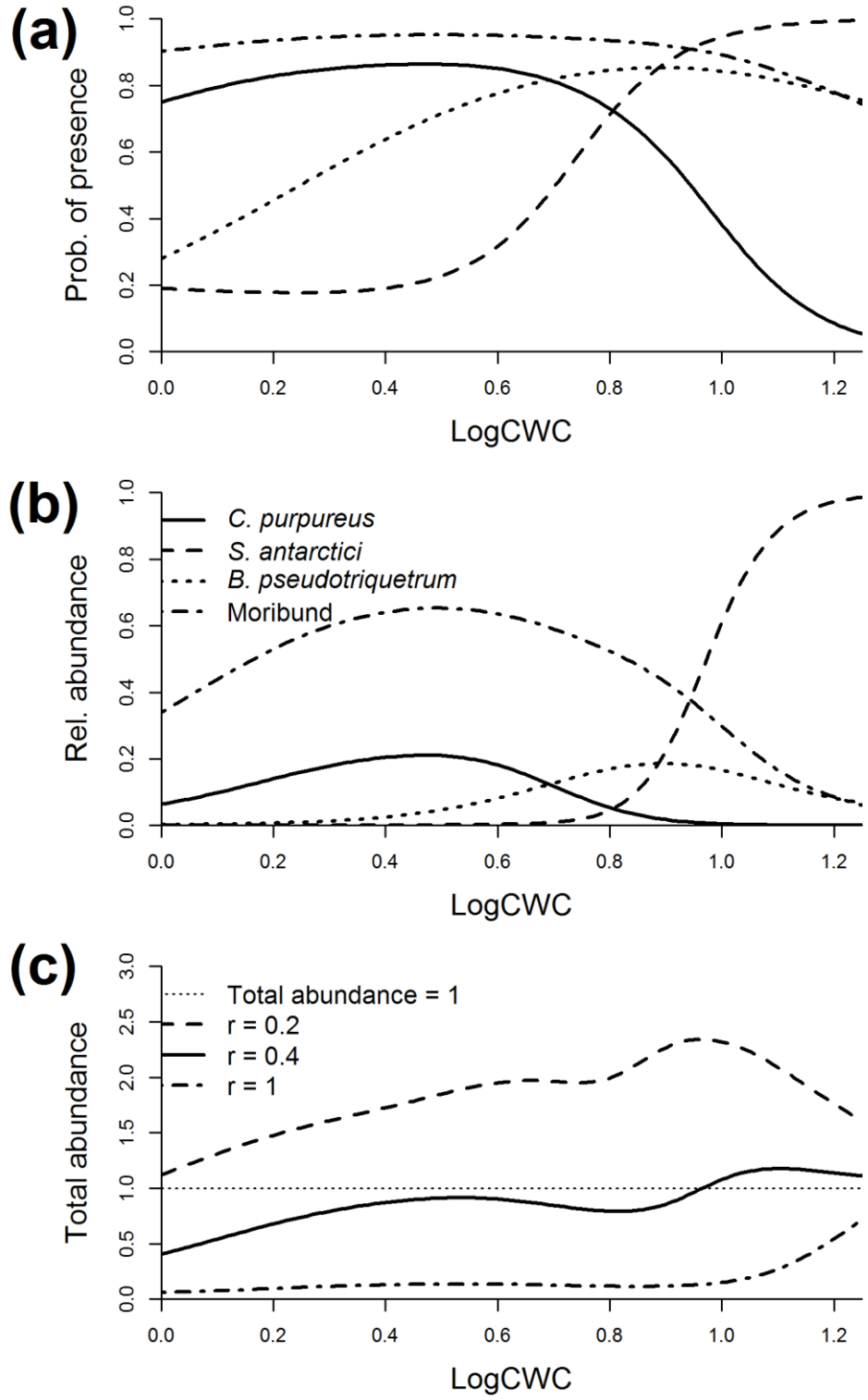
621


(a)
r = 0.4 N = 260000
mu = 260
mu = 2600
mu = 13000
mu = 26000
mu = 52000
mu = 130000
Probability
Rel. abund. of species in community (M / N)
(b)
r = 5 N = 260000
(c)
r = 0.4 N = 260000 n = 200
Rel. abund. of species in sample (m / n)
(d)
r = 5 N = 260000 n = 200
(e)
N = 260000 n = 200
r = 0.2
r = 0.4
r = 1
r = 5
Comm. rel. abund.(M / N)
Prob. pres. in sample (m > 0)
(f)
N = 1000 n = 200

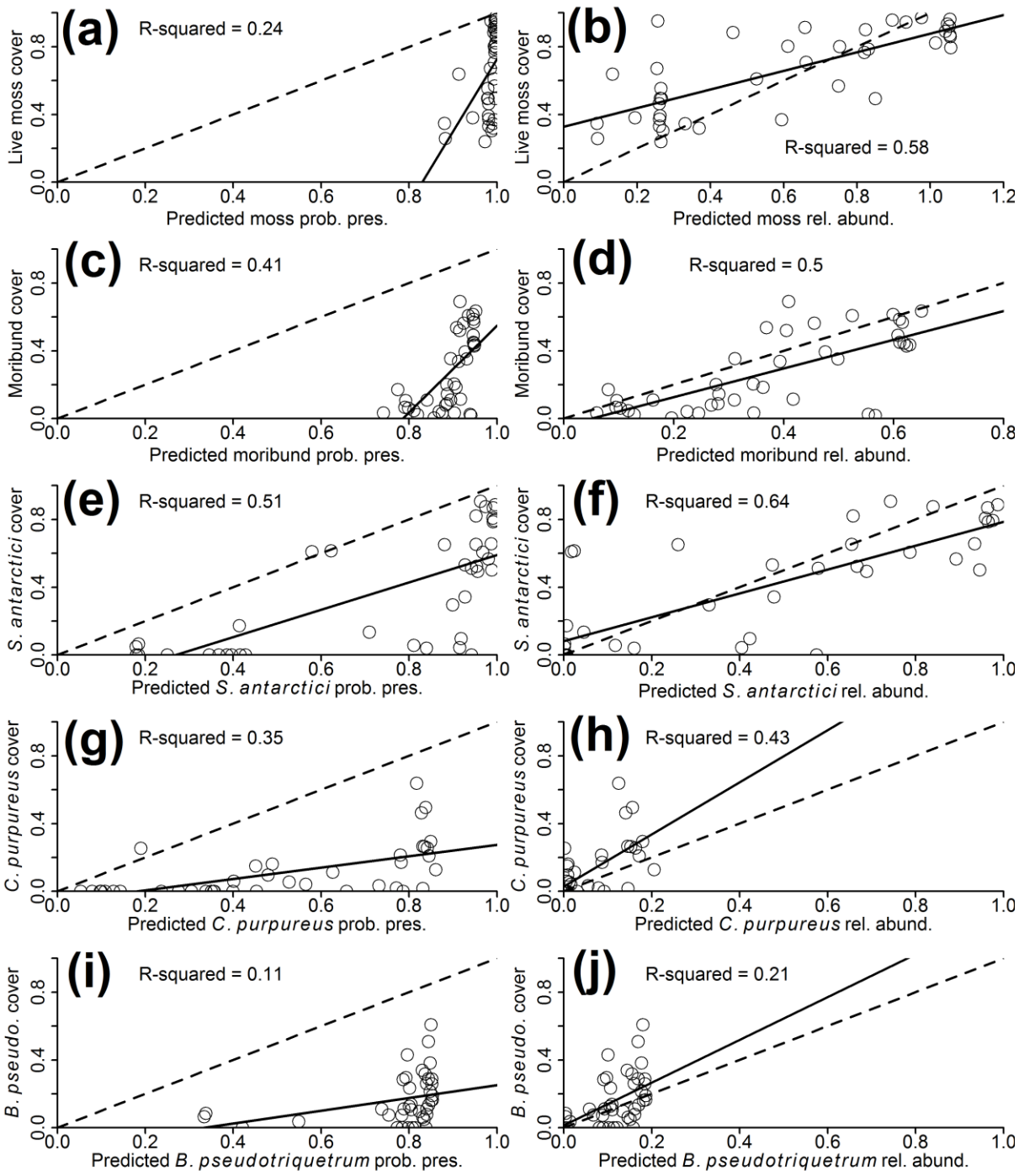
622

623

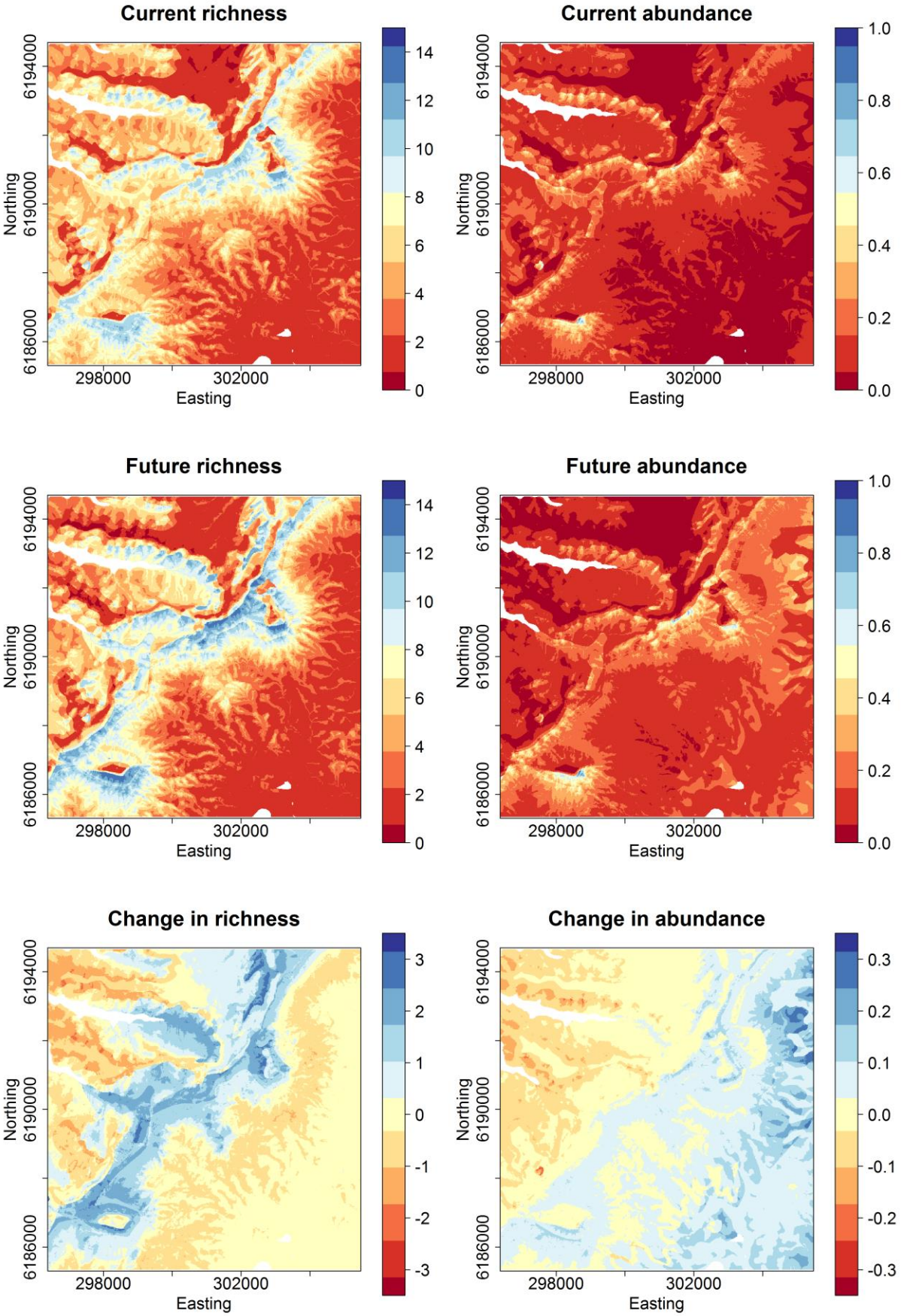
624


(a)
Prob. of presence
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
LogCWC
(b)
Rel. abundance
C. purpureus
S. antarctici
B. pseudotriquetrum
Moribund
LogCWC
(c)
Total abundance
Total abundance = 1
r = 0.2
r = 0.4
r = 1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
LogCWC

625


(a)
R-squared = 0.24
Live moss cover
Predicted moss prob. pres.
(b)
R-squared = 0.58
Live moss cover
Predicted moss rel. abund.
(c)
R-squared = 0.41
Moribund cover
Predicted moribund prob. pres.
(d)
R-squared = 0.5
Moribund cover
Predicted moribund rel. abund.
(e)
R-squared = 0.51
S. antarctici cover
Predicted S. antarctici prob. pres.
(f)
R-squared = 0.64
S. antarctici cover
Predicted S. antarctici rel. abund.
(g)
R-squared = 0.35
C. purpureus cover
Predicted C. purpureus prob. pres.
(h)
R-squared = 0.43
C. purpureus cover
Predicted C. purpureus rel. abund.
(i)
R-squared = 0.11
B. pseudo. cover
Predicted B. pseudotriquetrum prob. pres.
(j)
R-squared = 0.21
B. pseudo. cover
Predicted B. pseudotriquetrum rel. abund.

626


Current richness
Current abundance
Future richness
Future abundance
Change in richness
Change in abundance
Northing
Easting
6194000
6190000
6186000
298000
302000
0
2
4
6
8
10
12
14
0.0
0.2
0.4
0.6
0.8
1.0
-3
-2
-1
0
1
2
3
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3

627